

Technical Notes for Math, Theory, and Econometrics

Kaida Zhang

December 3, 2019

Part I

Statistics

Chapter 1

Basics

1.1 Distributions

Question 1.1.1. The 1st-order statistic among 10 i.i.d. obs and the 10th-order statistic among 100 i.i.d. obs, which one has larger mean?

Answer: The 1st-order among 10 has larger mean. Proof to be added. □

Theorem 1.1.1 (Law of Iterated Expectation). *We have:*

$$E[A] = E_B[E[A \mid B]]$$

and can be generalized to:

$$E[A \mid B] = E_{C|B}[E[A \mid B, C] \mid C]$$

Part II

Causal Inference

Chapter 2

Matching

2.1 General Matching

Glossary

ATT: Average Treatment on Treat group

ATC: Average Treatment on control group

ATE: Average Treatment Effect (on population)

The general idea for matching is that after controlling for some observable variables X , the expected Y_{0i} and Y_{1i} shall be the same on both the treatment and control group. Notice that for treatment group, $E[Y_{0i}]$ is counterfactual, and vice versa for control group.

Assumption 2.1.1 (Conditional Mean Independence (CMI)).

$$E[Y_{0i}|D_i = 1, X_i] = E[Y_{0i}|D_i = 0, X_i]$$

$$E[Y_{1i}|D_i = 1, X_i] = E[Y_{1i}|D_i = 0, X_i]$$

Obviously, the CMI is weaker than CIA (Conditional Independence). CMI only requires the equivalence of *expectation of potential outcome*, while CIA requires the equivalence of actual potential outcome.

We can form ATT:

$$\begin{aligned} ATT &\equiv E[Y_{1i} - Y_{0i}|D_i = 1] \\ &= E_{X_i|D_i=1} \{E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 1, X_i]\} \\ &= E_{X_i|D_i=1} \{E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 0, X_i]\} \\ &= E[Y_i|D_i = 1] - E_{X_i|D_i=1} \{E[Y_{0i}|D_i = 0, X_i]\} \end{aligned}$$

The last expression are taken expectation over distribution $F(X_i|D_i = 1)$, which is identified in the data. Line 2 to line 3 relies on the *assumption Conditional Mean Independence (CMI)*. Notice that each part is identified, thus ATT is identified as long as CMI is satisfied.

And similarly ATC:

$$\begin{aligned} ATC &\equiv E[Y_{1i} - Y_{0i}|D_i = 0] \\ &= E_{X_i|D_i=0} \{E[Y_{1i}|D_i = 0, X_i] - E[Y_{0i}|D_i = 0, X_i]\} \\ &= E_{X_i|D_i=0} \{E[Y_{1i}|D_i = 1, X_i] - E[Y_{0i}|D_i = 0, X_i]\} \\ &= E_{X_i|D_i=0} \{E[Y_{1i}|D_i = 1, X_i]\} - E[Y_i|D_i = 0] \end{aligned}$$

Since ATT and ATC are identified, it's natural that ATE are also identified:

$$ATE = \Pr(D_i = 1) \cdot ATT + \Pr(D_i = 0) \cdot ATC$$

Note that the treatment probability is unconditional and thus can be easily estimated from data.

But to **implement** the above estimand, we would need to estimate components like $E_{X_i|D_i=1}\{E[Y_{0i}|D_i=0, X_i]\}$, which has to be estimated non-parametrically.

As a first step, suppose X_i is discrete. Then:

$$E_{X_i|D_i=1}\{E[Y_{0i}|D_i=0, X_i]\} = \sum_x E[Y_{0i}|D_i=0, X_i=x]p(x|D_i=1)$$

where we can estimate $p(x|D_i=1)$ as:

$$\hat{p}(x|D_i=1) = \frac{\sum_{i=1}^N 1(X_i=x, D_i=1)}{\sum_{i=1}^N 1(D_i=1)}$$

and the estimator for ATT is given by:

$$\widehat{ATT} = \sum_x \left[\frac{\sum_{i=1}^N 1(X_i=x, D_i=0)Y_i}{\sum_{i=1}^N 1(X_i=x, D_i=0)} \cdot \hat{p}(x|D_i=1) \right]$$

With continuous X , the exact match cannot be executed and we would need to use some Kernel smooth. **To be added.**

2.2 Propensity Score

The core idea of propensity score is to use appropriate weights to in the control units to reproduce the covariates distribution in the treatment group. And we can use a single dimension score to summarize the selection bias. If we have enough data and all covariates are discrete, then we can simply calculate the propensity score by grouping data into stratum and simply counting the conditional probability of being treated in each stratum. If we have continuous covariates, then we need some model (typically a logit or probit) to estimate the propensity score.

The propensity score would be sufficient to estimate both ATT and ATC. By simple Bayes rule we can calculate $Pr(X_i=x|D_i=1)$:

$$Pr(X_i=x|D_i=1) = \frac{\Pr(D_i=1|X_i=x)\Pr(X_i=x)}{\Pr(D_i=1)}$$

then we can form \widehat{ATT} as in previous section.

Theorem 2.2.1 (Propensity Score Theorem). *If CIA is satisfied, i.e. $\{Y_{0i}, Y_{1i}\} \perp D_i | X_i$, then $\{Y_{0i}, Y_{1i}\} \perp D_i | P(X_i)$. Here $P(X_i) \equiv \Pr(D_i=1|X_i)$*

Proof. To be added. MHE p.81. □

Though we can use Bayes rule and the stratum to estimate ATE, this way is actually tedious. A easier way below, notice that: ¹

$$\begin{aligned} E\left[\frac{Y_i D_i}{p(X_i)}\right] &= E_{X_i}\{E\left[\frac{Y_i D_i}{p(X_i)}|X_i\right]\} \\ &= E_{X_i}\{E\left[\frac{Y_i \cdot 1}{p(X_i)}p(X_i) + \frac{Y_i \cdot 0}{p(X_i)}(1-p(X_i))\right]\} \\ &= E_{X_i}\{E[Y_i|D_i=1, X_i]\} \\ &= E_{X_i}\{E[Y_i|X_i]\} \\ &= E[Y_{1i}] \end{aligned}$$

1. Recall: $p(X_i) \equiv \Pr(D_i=1|X_i)$

Similarly, we have $E[\frac{Y_i(1-D_i)}{1-p(X_i)}] = E[Y_{0i}]$. Given propensity score $p(X_i)$, we can form an empirical estimator.

$$\begin{aligned} E[Y_{1i} - Y_{0i}] &= E[\frac{Y_i D_i}{p(X_i)}] - E[\frac{Y_i(1-D_i)}{1-p(X_i)}] \\ &= E[\frac{(D_i - p(X_i))Y_i}{p(X_i)(1-p(X_i))}] \end{aligned}$$

Simply replacing $E[\cdot]$ by $\frac{1}{N} \sum_{i=1}^N$ gives us the empirical counterpart for this estimator. This estimator also reveals the close relationship between regression estimator and matching estimator (below following MHE p.82).

Chapter 3

Instrument Variables

2SLS

The most basic instrument variable framework is:

$$y_i = \alpha + \beta X_i + \rho s_i + \varepsilon_i$$

we are interested in ρ while s_i is correlated with ε_i . But if we have some *exclusion variable* z_i which is uncorrelated with ε_i but is correlated with s_i , then we can form a first stage estimation:

$$s_i = \delta X_i + \gamma z_i + \eta_i$$

Plug into the initial problem:

$$y_i = \alpha + (\beta + \rho\delta)X_i + \rho\gamma z_i + (\rho\eta_i + \varepsilon_i)$$

this equation can be consistently estimated since all covariates are exogenous. Also note that the first stage also can be estimated. So we now can identify ρ :

$$\rho = \frac{\rho\gamma}{\gamma}$$

where the denominator is estimated from first stage and numerator from second stage.

Some important cases where z_i is a valid or invalid IV: (from Cyrus's lecture notes) The last case is especially interesting. Notice that in the last case we should *not* control over ε because that is a collider. Also the one with W as collider. So it actually does not matter how z and s are correlated, z does *not* have to have causal effect on s .

Wald Statistic

If the IV z_i is a binary variable, then 2SLS is Wald statistic. For Wald statistic:

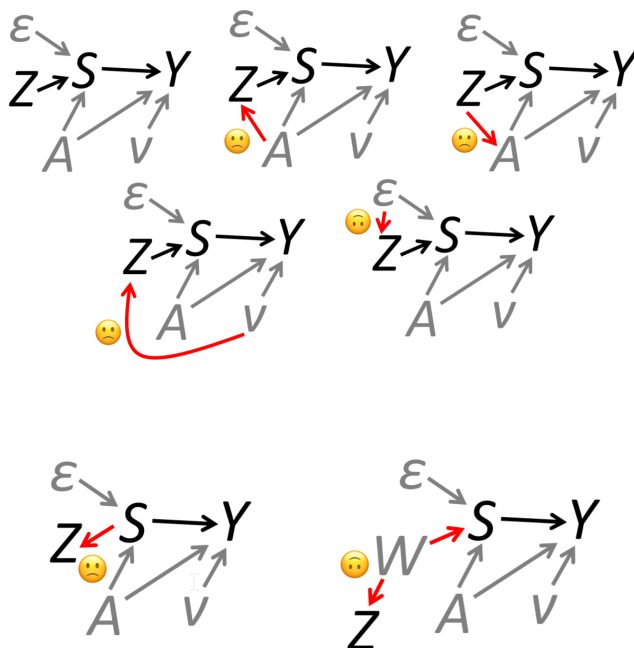
$$\rho = \frac{E[y_i|z_i = 1] - E[y_i|z_i = 0]}{E[s_i|z_i = 1] - E[s_i|z_i = 0]}$$

The intuition is straightforward, since the only channel for z_i to influence y_i is through s_i .

GMM Framework

We can also do IV in GMM framework. The benefit of using GMM is to have variance correct.

$$E[Z(Y - \beta X)] = 0$$



Chapter 4

Diff in Diff

The Diff-in-Diff estimators family include:

- Plain diff-in-diff
- Synthetic control. See Abadie, Diamond, and Hainmueller (2010), Xu (2017)

Chapter 5

Network Effect

Network in Migration:

- Bayer, Ross, and Topa (2008) leverages two effects: 1) the network effect is very local (say block); 2) the matching of unobservables happen at a larger neighborhood level. Then the migration difference of blocks within the same group can be attributed to network effects.
- Bryan A Stuart and Evan J Taylor (2019b, 2019a). They propose a new index to measure network effects, using individual level migration data.

5.1 Bayer, Ross, Topa, 2008 JPE

Patrick Bayer, Stephen L. Ross, and Giorgio Topa. 2008. “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes.” *Journal of Political Economy* 116, no. 6 (December): 1150–1196. ISSN: 0022-3808. doi:10.1086/595975

This paper is the first one that uses the geographic in local areas to identify the network effects. The identification assumption of their paper is:

1. The network effect is very local (say block level);
2. The matching of unobservables happen at a larger neighborhood level (say Census block group level)

Their baseline regression is:

$$W_{ij}^b = \rho_g + \alpha_0 R_{ij}^b + \varepsilon_{ij}$$

where i, j are two different individuals, W_{ij}^b is a dummy which equals 1 if i, j works in the same block; R_{ij}^b equals 1 if they reside in the same block; ρ_g is the fixed effect at block level. The parameter of interest is α_0

5.2 Stuart, Taylor, 2019 (wp)

Bryan A Stuart and Evan J Taylor. 2019b. “Migration Networks and Location Decisions: Evidence from U.S. Mass Migration” [in en]: 114

The basic idea of this paper is to use a diff-in-diff style estimator. Pick two similar born towns j and j' , which should have similar ex-ante migration probability to another town k . Then the observed migration difference D_{jk} and $D_{jk'}$

is explained by the network effect. They propose a simple index to capture the network effect, and they allow the network effect to differ across the destination towns.

The baseline regression is: (modified from Bayer, Ross, and Topa 2008)

$$D_{i,j(i),k} D_{i',j(i'),k} = \alpha_{g,k} + \sum_{j \in g} \beta_{j,k} 1[j(i) = j(i') = j] + \epsilon_{i,i',k}$$

where j is the birth town of i , k is the destination, $D_{i,j,k}$ is a dummy which equals 1 if i moves from j to k ; g is a group of towns where it is assumed that the ex-ante moving probability within the same group is the same. the parameter of interest is $\beta_{j,k}$. They show the following things:

1. $\beta_{j,k} = C_{j,k}$, where $C_{j,k} \equiv \sum_{i \neq i' \in j} \text{Cov}[D_{i,j,k}, D_{i',j,k}] / (N_j(N_j - 1))$ being the average covariance of the location decisions for two migrants from the same town;
2. $\alpha_{g,k} = P_{g,k}^2$, where $P_{j,k} \equiv \mathbb{E}[D_{i,j,k}]$ is the *ex-ante* probability of an individual migrating from j to k .

Their main argument: $\beta_{j,k}$ does not well capture the network effect, because it varies with ex-ante moving probability $P_{j,k}$:

$$\beta_{j,k} = C_{j,k} = P_{g,k}(\mu_{j,k} - P_{g,k})$$

where $\mu_{j,k} \equiv \mathbb{E}[D_{i,j,k} | D_{i',j,k} = 1]$.

To overcome this problem, they propose a intuitive network index $\Delta_{j,k}$, which captures the expected increase in the number of people from birth town j that move to destination k :

$$\Delta_{j,k} \equiv \mathbb{E}[N_{-i,j,k} | D_{i,j,k} = 1] - \mathbb{E}[N_{-i,j,k} | D_{i,j,k} = 0]$$

they show that $\Delta_{j,k}$ can be derived from observable moving patterns:

$$\Delta_{j,k} = \frac{(\mu_{j,k} - P_{g,k})(N_j - 1)}{1 - P_{g,k}} = \frac{C_{j,k}(N_j - 1)}{P_{g,k} - P_{g,k}^2}$$

Part III

IO

Chapter 6

Demand Estimation

6.1 Villas-Boas, 2007 RES

Sofia Berto Villas-Boas. 2007. “Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data.” *The Review of Economic Studies* 74, no. 2 (January): 625–652. doi:10.1111/j.1467-937x.2007.00433.x

The main goal of this paper is to recover the margins of both retailers and whosalers with only retailing data.

The demand side is typical BLP with random coefficients α_i, β_i :

$$U_{ijt} = d_j + d_t + x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \quad (1)$$

The paper focuses on the vertical structural of the supply side.

Scenario 1: simple linear pricing model The retailer profit function is:

$$\pi_{rt} = \sum_{j \in S_{rt}} [p_{jt} - p_{jt}^w - c_{jt}^r] s_{jt}(p)$$

and we can write FOC for product j of retailer r :

$$s_{jt} + \sum_{m \in S_{rt}} [p_{mt} - p_{mt}^w - c_{mt}^r] \frac{\partial s_{mt}}{\partial p_{jt}} = 0 \quad \forall j \in S_{rt}, \quad \text{for } r = 1, \dots, N_r \quad (5)$$

which is equivalent to the following equation holds for all commodities in the market \mathcal{J} :

$$s_{it} + \sum_{j \in \mathcal{J}} [p_{jt} - p_{jt}^w - c_{jt}^r] \mathbb{1}\{i, j \in \mathcal{J}_r\} \frac{\partial s_{jt}}{\partial p_{it}} = 0, \quad \forall i \in \mathcal{J}$$

Then stack all products in the market together in the market (this step is tricky):

$$p_t - p_t^w - c_t^r = -(T_r * \Delta_{rt})^{-1} s_t(p)$$

where T_r is the ownership matrix, $T_r(i, j)$ equals to 1 if i, j belong to the same owner, other wise 0. Δ_{rt} is the retailer's response matrix, $\Delta_{rt}(i, j) = \frac{\partial s_{jt}}{\partial p_{it}}$. Both p_t^w and c_t^r are unobservable, other parts are observable. Δ_{rt} is estimated from the demand side.

Similarly, for wholesaler we have:

$$\pi_{wt} = \sum_{j \in S_{wt}} [p_{jt}^w - c_{jt}^w] s_{jt}(p(p^w))$$

the corresponding FOC is:

$$s_{jt} + \sum_{m \in S_{wt}} [p_{mt}^w - c_{mt}^w] \frac{\partial s_{mt}}{\partial p_{jt}^w} = 0 \quad \forall j \in S_{wt}, \quad \text{for } w = 1, \dots, N_w$$

The $\frac{\partial s_{mt}}{\partial p_{jt}^w}$ for wholesalers are very complicated, because the wholesale price change in one product will induce all retailers to re-optimize their products. Let $\Delta_{wt}(i, j)$ be $\frac{\partial s_{mt}}{\partial p_{jt}^w}$. First notice that $\Delta_{wt} = \Delta'_{pt} \Delta_{rt}$, where Δ_{pt} is the derivative matrix of retail price to wholesale price. So the key for wholesaler FOC is to find Δ_{pt} . To obtain this, do total differentiation for all retail prices $dp_k, k = 1, \dots, N$ with respect to a wholesale price p_f^w in the optimization equation (5) of retailers:

$$\sum_{k=1}^N \left[\underbrace{\frac{\partial s_j}{\partial p_k} + \sum_{i=1}^N \left(T_r(i, j) \frac{\partial^2 s_i}{\partial p_j \partial p_k} (p_i - p_i^w - c_i^r) \right) + T_r(k, j) \frac{\partial s_k}{\partial p_j}}_{g(j, k)} dp_k - \underbrace{T_r(f, j) \frac{\partial s_f}{\partial p_j}}_{h(j, f)} dp_f^w = 0$$

Stack all products together, we have:

$$\frac{dp}{dp_f^w} = G^{-1} H_f$$

where G, H are matrixes for g, h respectively. Thus the FOC of wholesalers can be written as:

$$p_t^w - c_t^w = - (T_w * \Delta_{wt})^{-1} s_t(p)$$

Chapter 7

Dynamic Estimation

Rust (1987) proposes Fixed Nested Point approach. The bad thing is computation complexity.

7.1 Hendel, Nevo, 2006 EMCA

Igal Hendel and Aviv Nevo. 2006. “Measuring the Implications of Sales and Consumer Inventory Behavior” [in en]. *Econometrica* 74 (6): 1637–1673. ISSN: 1468-0262. doi:10.1111/j.1468-0262.2006.00721.x

This paper estimates a model with consumer storage. They solved two problems: 1) the inventory (a state variable) is unobservable to researchers; 2) the state space is too large.

A key feature of their model is that they can separate a static problem from the dynamic problem, and use that static problem to estimate the purchasing utility parameters, which includes the brand heterogeneity. The dynamic problem is then independent of brand, but solely quantity related.

Model The flow utility is assumed to be:

$$u(c_{ht}, \nu_{ht}; \theta_h) + \alpha_h m_{ht}$$

where c_{ht} is the quantity consumed (not purchased) by consumer h in time t ; m_{ht} is the consumption of the outside good. Note that c_{ht} is assumed to be the total consumption of all brands: $c_{ht} \equiv \sum_j c_{jht}$. This implies: 1) the consumption has no brand preference; 2) the consumption will not be affected by which brand is at storage.

Then we can write the dynamic problem:

$$\begin{aligned} V(s_1) = & \max_{\{c_h(s_t), d_{hj}(s_t)\}} \sum_{t=1}^{\infty} \delta^{t-1} E[u(c_{ht}, \nu_{ht}; \theta_h) - C_h(i_{h,t+1}; \theta_h) \\ & + \sum_j d_{hjxt} (\alpha_h p_{jxt} + \xi_{hjt} + \beta_h a_{jxt} + \varepsilon_{hjt}) | s_1] \\ \text{s.t.} \quad & 0 \leq i_{ht}, \quad 0 \leq c_{ht}, \quad 0 \leq x_{ht}, \quad \sum_{j,x} d_{hjxt} = 1 \\ & i_{h,t+1} = i_{ht} + x_{ht} - c_{ht} \end{aligned}$$

where i_{ht} is the inventory of consumer h at time t ; d_{hjxt} equals 1 if brand j and size x is purchased; C_h is the storage cost; a_{jht} is the advertising variable.

Notice, the purchasing per se (in addition to the following consumption) has utility here, which I do not understand. Also, the product differentiation takes place at *purchase*, not *consumption*. As mentioned by the authors, this

assumption “helps reduce the state space. Instead of the whole vector of brand inventories, only the total quantity in stock matters, regardless of brand.”

Estimation The estimation process of this dynamic problem is not trivial. First, the state variable inventory i_t is not observable. Second, the state space include brand, variety, price and advertisements, which is very large and intractable.

To solve the unobservable inventory problem, they use the fact that the optimal consumption can be calculated. For this moment assume that the inventory i_t is observable, then the researcher can derive the optimal consumption c_t since all state variable now observable. Then the researcher can derive the purchasing probability d_{t+1} which is observed in the data, and, therefore, derive i_{t+1} . The probability of the observed sequence of purchases is:

$$\begin{aligned} & \Pr(d_1 \cdots d_T | p_1 \cdots p_T) \\ &= \int \prod_{t=1}^T \Pr(d_t | p_t, i_t(d_{t-1}, \dots, d_1, \\ & \quad \nu_{t-1}, \dots, \nu_1, i_1), \nu_t) dF(\nu_1, \dots, \nu_T) dF(i_1) \end{aligned}$$

where $F(i_1)$ is observed in the data.

Dimension Reduction The dimension is too high for practical estimation. So the paper uses *inclusive value* to reduce the dimension. The steps are as following:

1. Estimate the static parameters, which are purchasing utility parameters;
2. Simplify the state of DP using inclusive value;
3. Use nested fixed point to estimated the simplified DP problem.

Recall that the consumption utility does not depend on the brand but only on the quantity; then it is natural that the optimal consumption depend only on the quantity of storage, independent of the brand. (details see their lemma 1) This is an important assumption for them to separate the consumption and the purchasing. Then it's convenient to define:

$$M(s_t, j, x) \equiv \text{Max}_c \{u(c + \nu_t) - C(i_{t+1}) + \delta E(V(s_{t+1}) | d_{jx}, c, s_t)\}$$

The above paragraph is saying that, conditional on quantity x , $M(s_t, j, x) = M(s_t, j', x)$. Then it's direct to show that:

$$\begin{aligned} \Pr(d_{jt} = 1 | x_t, i_t, p_t, \nu_t) &= \frac{\exp(\alpha p_{jxt} + \xi_{kx} + \beta a_{kxt})}{\sum_k \exp(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt})} \\ &= \Pr(d_{jt} = 1 | x_t, p_t) \end{aligned}$$

Using conditional probability rule, we can rewrite the joint choice probability of j and x to be:

$$\begin{aligned} \Pr(d_{jt} = 1, x_t | p_t, i_t, \nu_t) &= \Pr(d_{jt} = 1 | p_t, x_t, i_t, \nu_t) \Pr(x_t | p_t, i_t, \nu_t) \\ &= \Pr(d_{jt} = 1 | x_t, p_t) \Pr(d_{jt} = 1 | p_t, x_t, i_t, \nu_t) \end{aligned}$$

In the first step, we estimate the first part separately, which give us purchasing utility parameters.

Then in the second step, we show that the DP problem can be simplified using the inclusive value. The inclusive value ω_{xt} is defined as:

$$\omega_{xt} = \log \left\{ \sum_k \exp(\alpha p_{kxt} + \xi_{kx} + \beta a_{kxt}) \right\}$$

Notice that the ω_{xt} is known up to parameters which can be estimated in the first step. This simplifies the DP problem because ω_{xt} is a scalar, so we do not have to track the price distribution for different brands. So if we can rewrite the DP using ω_{xt} , we solve the problem of It can be shown that the DP problem is equivalent to the below DP using the inclusive value: (see their appendix for proof)

$$\begin{aligned} & V(i_t, \omega_t, \varepsilon_t, \nu_t) \\ &= \text{Max}_{\{c, x\}} \{u(c + \nu_t) - C(i_{t+1}) + \omega_{xt} + \varepsilon_{xt} \\ & \quad + \delta E[V(i_{t+1}, \omega_{t+1}, \varepsilon_{t+1}, \nu_{t+1}) | i_t, \omega_t, \varepsilon_t, \nu_t, c, x]\} \end{aligned}$$

The key of the proof is to use the fact that all error terms are i.i.d. and T1EV distributed, so there is a closed form of expected maximum utility over alternative brands, which is ω_{xt} .

With this, we can calculate the probability of a particular purchasing sequence using previous chain method.

Chapter 8

Observational Learning

Empirical Observational Learning:

- Literature Review: Sorensen (2017)
- IO: Moretti (2011), Hendricks, Sorensen, and Wiseman (2012), Newberry (2016), Li and Wu (2019), Hollenbeck, Moorthy, and Proserpio (2019)
- Development: Bobonis and Finan (2009), Angelucci et al. (2010), Bursztyn et al. (2014)

The *observational learning* is often frequently associated with *peer effects*. Observational learning can be seen as a mechanism underlying peer effect (Bursztyn et al. 2014).

Part IV

Education

Chapter 9

Demand for Education

A huge literature in school choice:

- Theory paper: Barseghyan, Clark, and Coate (2019)
- Empirical paper: Hastings, Kane, and Staiger (2009), Deming, Hastings, Kane, and Staiger (2014), Burgess, Greaves, Vignoles, and Wilson (2015), Walters (2018)
- Combine with location choice:
 - Methodology paper combining residential sorting with school choice: Rothstein (2006), Bayer and Timmins (2007), Bayer, Ferreira, and McMillan (2007), Bayer and McMillan (2010), Bayer et al. (2016)
 - Empirical: Cordes, Schwartz, and Stiefel (2019) uses NYC residential data to study the effect of mobility; Epple, Jha, and Sieg (2018) seems to also estimate a demand model with student sorting

Some other papers focus more on the school competition:

- Reduced form: Lavy (2010) uses DiD and border discontinuity to show that the school competition in Tel-Aviv has positive results.

There are certain reasons why the competition among schools will not work.

1. Schools (principals and teachers) may lack the incentive to correspond to the competition
2. Schools may lack the flexibility to change. E.g. they cannot fire under-qualified teachers at their own discretion.
3. Families may lack incentive to correspond to the “value added” of schools, but simply prefer schools with better peers (Abdulkadiroglu, Pathak, and Schellenberg 2017; W Bentley MacLeod and Urquiola 2015, 2019).
4. Families may not be able to identify effective schools (Allende, Gallego, and Neilson 2018)

Heterogeneous student response to inputs:

- Bau (2019) shows that rich students are more responsive to quality.
- Jackson, Johnson, and Persico (2016) shows that school inputs are more valuable to poor students (p.192)

Endogenous school differentiation:

- W Bentley MacLeod and Urquiola (2012), Gilraine, Petronijevic, and Singleton (2019)

Review: Demand for schools

Hastings, Kane, and Staiger (2009) They use a simple exploded logit model to show that the demand for high academic quality is heterogeneous across families. The more affluent families put higher weights on academic performance, but poorer families and minorities put higher weights in the race composition of the school.

Cons:

1. In their utility function (4), they directly include school average test score S_j , which is obviously endogenous of student composition. They do not deal with this using IV.
2. They do not have an unobserved attractiveness term ξ_j as BLP in the utility.

Epple, Jha, and Sieg (2018) The main contribution of this paper is that they incorporate the peer effects into demand estimation. They also deal with the limited capacity problem by introducing a *shadow price* for over-subscribed schools. Note that this is actually tricky, because when they estimate the demand, they assume no capacity constraint, and the constraint only holds in computing the counterfactual equilibrium.

The main contribution of their paper is an algorithm of calculating the counterfactual equilibrium with peer effects. The demand estimation part is quite standard. They use MLE for the demand estimation.

Cons:

1. The demand estimation does not consider the capacity constraint
2. The location sorting is not considered in the model

9.1 Review: School Competition

Charter School Spillover Effects

Booker, Gilpatric, Gronberg, and Jansen (2008) This paper uses a longitudinal data in Texas to show that charter schools in Texas have positive spillovers in nearby TPS. They use panel data to control student/campus fixed effect to exclude the sorting effect. They focus on the response at district level instead of campus level. One charter school nearby increase the Math and Reading test score by 0.021 sd; as a comparison, those who move to another campus have a negative effect of 0.104 sd on Math. They also find the effect is more salient for black and Hispanic students.

9.2 Bayer, Timmins, 2007 EJ

Patrick Bayer and Christopher Timmins. 2007. “Estimating Equilibrium Models of Sorting Across Locations” [in en]. *The Economic Journal* 117, no. 518 (March): 353–374. ISSN: 0013-0133, 1468-0297. doi:10.1111/j.1468-0297.2007.02021.x

This paper proposes a way to estimate the local spillover effects (summarized by the share of people living in location j , σ_j). The basic idea is to use the BLP framework, but have σ_j in the utility function. They show that there is no need to exactly solve the model, and the complication mainly comes from multi-equilibria phenomena which they solve by proposing a semi-likelihood estimator as in BBL. They do not encompass the followings in the paper: the endogenous price, the school choice problem. But both should be not difficult to deal with.

The utility is:

$$U_{i,j} = \mathbf{X}_j' \beta_i + \alpha_i \sigma_j + \xi_j + \varepsilon_{i,j} \quad (9.1)$$

Note:

- σ_j is share of individuals who choose location j . This is definitely endogenous since it's determined in the equilibrium, and needs to be instrumented.
- There is no price included in above utility. If included, needs additional instrument for price.
- Other parts are standard BLP

Equilibrium:

Given σ, ξ (both being vector), the individual choice probability is:

$$P_{i,j} = g_{i,j}(\mathbf{Z}_i, \mathbf{X}, \sigma, \xi) \forall i, j$$

where we have the closed form for $g(\cdot)$ if $\varepsilon_{i,j}$ follows i.i.d. T1EV.

Aggregate over individuals and we get the shares:

$$\sigma_j = \int g_{i,j}(\mathbf{Z}_i, \mathbf{X}, \sigma, \xi) dh(\mathbf{Z}) \forall j$$

where $h(Z)$ is the density of Z in the population, which is known. KZ: notice that we can use Nevo strategy if we can observe individual level demographics.

Then the above two equations define the equilibrium. The uniqueness is not guaranteed in general, unless the local spillover effects α_i are constrained to be negative. See Bayer and McMillan (2010) for more details on uniqueness conditions.

Estimation:

Sketch:

- The estimation is basically BLP, but we need to find instruments for market shares σ_j . The two steps approach below is very similar to that in Hackmann (2019).
- The equilibrium may have multiple equilibria, needs to be dealt with. The authors pick an equilibrium selection rule $P(\sigma|\Omega)$ and non-parametrically estimate this probability along with the likelihood. But other more robust approach may be used as well, like the moment inequality in Ciliberto and Tamer (2009).
- Step 1: Estimate δ_j as fixed effects, as well as β_1, α_1
- Step 2: Estimate the mean utility parameters within δ_j

First rewrite the utility function to be the addition of mean utility δ_j and idiosyncratic utility part:

$$U_{i,j} = \delta_j + \mathbf{X}'_j \mathbf{Z}_i \beta_1 + \sigma_j \mathbf{Z}_i \alpha_1 + \varepsilon_{i,j}$$

where

$$\delta_j = \mathbf{X}'_j \beta_0 + \alpha_0 \sigma_j + \xi_j$$

Since the possibility of multiple equilibria, the likelihood is not well defined. The authors then define $P(\sigma|\Omega)$ as the equilibrium selection rule, where Ω contains exogenous data $\{X, Z\}$ and parameters $\{\delta, \alpha_1, \beta_1, \Sigma\}$. Importantly, Ω does not contain unobservable individual level shocks ε_{ij} .

$$L = P(\sigma|\Omega) \prod_i \prod_j P_i(j|\sigma, \Omega)^{I_{ij}}$$

Here, $P(\sigma|\Omega)$ does not need to be estimated! (see later); and for given (σ, Ω) , the $P_i(j)$ can be calculated. **Importantly**, different from standard entry model in IO like Ciliberto and Tamer (2009), here our equilibrium selection rule does not

depend on unobservable terms ε_{ij} . This relies on the assumption that for all observable $\{X, Z\}$, there are continuum number of consumers with different ε_{ij} . This is similar to the incomplete information assumption made in Seim (2006). Therefore, the likelihood of the two components are orthogonal to each other, and we can estimate the second part only, assuming that each individual behaves optimally in the *observed* equilibrium.

Instruments:

The instruments are the most important contribution of this paper.

In the first step MLE, where the mean utility δ_j is estimated as a fixed effect, we do not need IV. But we use GMM to estimate the parameters within mean utility, then we need IV. We need IV because σ_j is correlated with ξ_j . So we need IV which is correlated with σ_j but uncorrelated with ξ_j .

The IV chosen here are similar to BLP IV, they choose the fixed attributes of other (nearby) locations. They further uses a function to summarize the impact of such fixed attributes to a single instrument of share. “This corresponds to using the predicted share of individuals that chooses a location based only on the observed, exogenous choice and individual characteristics used in the model, ignoring the role of local spillovers. ”

For the identification to really work, one needs some choice set variations. For example, there are several markets and in each market the consumers will have different choice sets. Their table 3 shows that IV approach performs poorly without choice set variation.

9.3 Bayer, McMillan, 2010 wp

Patrick Bayer and Robert McMillan. 2010. *Choice and Competition in Local Education Markets* [in en]. Technical report

The main purpose of this paper is to investigate whether the schools will respond to the competition.

$$T^m = X_c^m \gamma_c + X_n^m \gamma_n + \gamma_E E^m + \varepsilon^m$$

where T^m is the average test score for school m , X_c for household characteristics, X_n for neighborhood controls, E^m is the slope for demand-quantity curve. The parameter of interest is γ_E . Assume that the only unobservable is the school and teacher effort. Think of the two polarized cases: 1) the school only cares about its teaching quality; 2) the school is rent seeking. If the school only cares about teaching quality, then the schools’ production function shall not be related with the competition, but only related with the observables and we can control that. **This claim is tricky.** But if the school is rent seeking, then the schools face the following trade-offs: raising quality (by raising effort) can make the school more attractive which is valuable, but the induced higher effort is also costly. Whether increasing quality is worthwhile depends on the responsiveness of demand to quality, the E^m .

Since E^m is not observed directly in the data, it needs to be estimated. They use **housing price** as the proxy for demand and use **school average test scores** as proxy for quality, conditional on school and student characteristics. **KZ: notice this quality is “net quality”, trying to capture the effort induced quality.** In order to overcome the endogeneity, they construct a discrete choice model, simultaneously choosing schools and locations, using BLP and methods in Bayer, Ferreira, and McMillan (2007). **Learn this technique.** Then they conduct the counterfactual by increasing the test score of school m by 1 s.d., then re-estimate the model and see how much the housing price will increase. The resulting change captures the responsiveness of demand to quality.

Chapter 10

School Finance

Jackson, Johnson, and Persico (2016) uses the School Finance Reforms (SFR) as exogenous variations of funding levels across school districts. They use this variation to conduct DiD to show that higher expenditure will improve the student performance.

Bibliography

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program” [in en]. *Journal of the American Statistical Association* 105, no. 490 (June): 493–505. ISSN: 0162-1459, 1537-274X. doi:10.1198/jasa.2009.ap08746.
- Abdulkadiroglu, A, P A Pathak, and J Schellenberg. 2017. “Do Parents Value School Effectiveness?” (January).
- Allende, Claudia, Francisco Gallego, and Christopher Neilson. 2018. “Approximating the Equilibrium Effects of Informed School Choice” (December): 1–58.
- Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel, and Imran Rasul. 2010. “Family Networks and School Enrolment: Evidence from a Randomized Social Experiment.” *Journal of Public Economics* 94, no. 3 (April): 197–221. ISSN: 0047-2727. doi:10.1016/j.jpubeco.2009.12.002.
- Barseghyan, Levon, Damon Clark, and Stephen Coate. 2019. “Peer Preferences, School Competition, and the Effects of Public School Choice” [in en]. *American Economic Journal: Economic Policy* 11, no. 4 (November): 124–158. ISSN: 1945-7731, 1945-774X. doi:10.1257/pol.20170484.
- Bau, Natalie. 2019. “School Competition and Product Differentiation” (August): 1–60.
- Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. “A Unified Framework for Measuring Preferences for Schools and Neighborhoods” [in en]. *Journal of Political Economy* 115, no. 4 (August): 588–638. ISSN: 0022-3808, 1537-534X. doi:10.1086/522381.
- Bayer, Patrick, and Robert McMillan. 2010. *Choice and Competition in Local Education Markets* [in en]. Technical report.
- Bayer, Patrick, Robert McMillan, Alvin Murphy, and Christopher Timmins. 2016. “A Dynamic Model of Demand for Houses and Neighborhoods.” *Econometrica* 84, no. 3 (May): 893–942. ISSN: 0012-9682. doi:10.3982/ECTA10170.
- Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes.” *Journal of Political Economy* 116, no. 6 (December): 1150–1196. ISSN: 0022-3808. doi:10.1086/595975.
- Bayer, Patrick, and Christopher Timmins. 2007. “Estimating Equilibrium Models of Sorting Across Locations” [in en]. *The Economic Journal* 117, no. 518 (March): 353–374. ISSN: 0013-0133, 1468-0297. doi:10.1111/j.1468-0297.2007.02021.x.
- Bobonis, Gustavo J, and Frederico Finan. 2009. “Neighborhood Peer Effects in Secondary School Enrollment Decisions.” *The Review of Economics and Statistics* 91, no. 4 (October): 695–716. ISSN: 0034-6535. doi:10.1162/rest.91.4.695.
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. “The Effect of Charter Schools on Traditional Public School Students in Texas: Are Children Who Stay behind Left Behind?” *Journal of Urban Economics* 64, no. 1 (July): 123–145. ISSN: 0094-1190. doi:10.1016/j.jue.2007.10.003.
- Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson. 2015. “What Parents Want: School Preferences and School Choice” [in en]. *The Economic Journal* 125, no. 587 (September): 1262–1289. ISSN: 00130133. doi:10.1111/eoj.12153.

- Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions." *Econometrica* 82, no. 4 (July): 1273–1301. ISSN: 0012-9682. doi:10.3982/ECTA11991.
- Ciliberto, Federico, and Elie Tamer. 2009. "Market Structure and Multiple Equilibria in Airline Markets." *Econometrica* 77, no. 6 (January): 1791–1828. doi:10.3982/ECTA5368.
- Cordes, Sarah A., Amy Ellen Schwartz, and Leanna Stiefel. 2019. "The Effect of Residential Mobility on Student Performance: Evidence From New York City" [in en]. *American Educational Research Journal* 56, no. 4 (August): 1380–1411. ISSN: 0002-8312. doi:10.3102/0002831218822828.
- Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. 2014. "School Choice, School Quality, and Postsecondary Attainment" [in en]. *American Economic Review* 104, no. 3 (March): 991–1013. ISSN: 0002-8282. doi:10.1257/aer.104.3.991.
- Epple, Dennis, Akshaya Jha, and Holger Sieg. 2018. "The Superintendent's Dilemma: Managing School District Capacity as Parents Vote with Their Feet." *Quantitative Economics* 9, no. 1 (April): 483–520. doi:10.3982/QE592.
- Gilraine, Michael, Uros Petronijevic, and John D Singleton. 2019. "Horizontal Differentiation and the Policy Effect of Charter Schools" [in en]: 60.
- Hackmann, Martin B. 2019. "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry" [in en]. *American Economic Review* 109, no. 5 (May): 1684–1716. ISSN: 0002-8282. doi:10.1257/aer.20151057.
- Hastings, Justine, Thomas J Kane, and Douglas Staiger. 2009. "Heterogeneous Preferences and the Efficacy of Public School Choice" [in en]: 62.
- Hendel, Igal, and Aviv Nevo. 2006. "Measuring the Implications of Sales and Consumer Inventory Behavior" [in en]. *Econometrica* 74 (6): 1637–1673. ISSN: 1468-0262. doi:10.1111/j.1468-0262.2006.00721.x.
- Hendricks, Kenneth, Alan Sorensen, and Thomas Wiseman. 2012. "Observational Learning and Demand for Search Goods" [in en]. *American Economic Journal: Microeconomics* 4, no. 1 (February): 1–31. ISSN: 1945-7669, 1945-7685. doi:10.1257/mic.4.1.1.
- Hollenbeck, Brett, Sridhar Moorthy, and Davide Proserpio. 2019. "Advertising Strategy in the Presence of Reviews: An Empirical Analysis" [in en]. *Marketing Science* (September): mksc.2019.1180. ISSN: 0732-2399, 1526-548X. doi:10.1287/mksc.2019.1180.
- Jackson, C. Kirabo, Rucker C. Johnson, and Claudia Persico. 2016. "The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms" [in en]. *The Quarterly Journal of Economics* 131, no. 1 (February): 157–218. ISSN: 0033-5533. doi:10.1093/qje/qjv036.
- Lavy, Victor. 2010. "Effects of Free Choice Among Public Schools" [in en]. *The Review of Economic Studies* 77, no. 3 (July): 1164–1191. ISSN: 0034-6527. doi:10.1111/j.1467-937X.2009.00588.x.
- Li, Xitong, and Lynn Wu. 2019. "Herding and Social Media Word-of-Mouth: Evidence from Groupon" [in en]: 49.
- MacLeod, W Bentley, and Miguel Urquiola. 2012. "Competition and Educational Productivity: Incentives Writ Large" [in en]: 38.
- . 2015. "Reputation and School Competition." *American Economic Review* 105, no. 11 (November): 3471–3488. doi:10.1257/aer.20130332.
- . 2019. "Is Education Consumption or Investment? Implications for School Competition." *Annual Review of Economics* 11 (1): 563–589. doi:10.1146/annurev-economics-080218-030402.
- Moretti, Enrico. 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales" [in en]. *The Review of Economic Studies* 78, no. 1 (January): 356–393. ISSN: 0034-6527. doi:10.1093/restud/rdq014.
- Newberry, Peter W. 2016. "An Empirical Study of Observational Learning" [in en]. *The RAND Journal of Economics* 47, no. 2 (May): 394–432. ISSN: 07416261. doi:10.1111/1756-2171.12132.
- Rothstein, Jesse M. 2006. "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions" [in en]. *THE AMERICAN ECONOMIC REVIEW* 96 (4): 23.

- Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55 (5): 999–1033. ISSN: 0012-9682. doi:10.2307/1911259.
- Seim, Katja. 2006. "An Empirical Model of Firm Entry with Endogenous Product-type Choices." *The RAND Journal of Economics* 37, no. 3 (September): 619–640. doi:10.1111/j.1756-2171.2006.tb00034.x.
- Sorensen, Alan T. 2017. "Bestseller Lists and the Economics of Product Discovery." *Annual Review of Economics* 9 (1): 87–101. doi:10.1146/annurev-economics-080614-115708.
- Stuart, Bryan A., and Evan J. Taylor. 2019a. "The Effect of Social Connectedness on Crime: Evidence from the Great Migration" [in en]. *The Review of Economics and Statistics* (August): 1–46. ISSN: 0034-6535, 1530-9142. doi:10.1162/rest_a_00860.
- . 2019b. "Migration Networks and Location Decisions: Evidence from U.S. Mass Migration" [in en]: 114.
- Villas-Boas, Sofia Berto. 2007. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *The Review of Economic Studies* 74, no. 2 (January): 625–652. doi:10.1111/j.1467-937x.2007.00433.x.
- Walters, Christopher R. 2018. "The Demand for Effective Charter Schools." *Journal of Political Economy* 126, no. 6 (December): 2179–2223. doi:10.1086/699980.
- Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models" [in en]. *Political Analysis* 25, no. 1 (January): 57–76. ISSN: 1047-1987, 1476-4989. doi:10.1017/pan.2016.2.